CURRENT CONCEPTS REVIEW
USER'S GUIDE TO THE ORTHOPAEDIC LITERATURE: HOW TO USE AN ARTICLE ABOUT PROGNOSIS

BY MOHIT BHANDARI, MD, MSC, GORDON H. GUYATT, MD, MSC, AND MARC F. SWIONTKOWSKI, MD

Investigation performed at the Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada, and the Department of Orthopaedic Surgery, University of Minnesota, Minneapolis, Minnesota

---

➤ Prognosis studies are investigations examining the possible outcomes of a disease or operative procedure and the probability with which they can be expected to occur.

➤ Primary guides for assessing the validity (study methodology) of a prognosis study are:
- Was there a representative sample of patients?
- Were the patients sufficiently homogeneous with respect to prognostic risk? If not, did the investigators provide estimates for all clinically relevant subgroups?

➤ Secondary guides for assessing the validity (study methodology) of a prognosis study are:
- Was follow-up sufficiently complete?
- Were objective and unbiased outcome criteria used?

Clinical Scenario

You are an orthopaedic surgeon consulting on the case of a seventy-seven-year-old woman with osteoarthritis in the right hip causing pain and functional impairment who was referred to you by a local family physician. The woman had a left total hip arthroplasty twelve years ago, with a good result. For the present problem, she received a course of conservative therapy including anti-inflammatory medications and physiotherapy. She currently uses a cane to walk and is no longer able to do housework.

On examination, she is found to be moderately overweight (67 kg) and 5 ft (152.4 cm) tall. She has a 2-cm limb-length discrepancy and a severely restricted range of motion of the right hip. Examination of anteroposterior radiographs of the pelvis and the right hip reveal advanced osteoarthritis with large osteophytes, subchondral cysts, and decreased joint space. Additional evaluation of the radiograph of the right hip reveals a femoral canal-flare index (the canal width 20 mm proximal to the geometric center of the lesser trochanter divided by the canal width at the isthmus of the femur) of 2.0. Evaluation of a radiograph of the left hip, in which a so-called hybrid hip arthroplasty (fixation of the acetabular component without cement and the femoral component with cement) was done, reveals no radiographic evidence of loosening.

A total hip arthroplasty of the right hip is recommended. The patient seems willing to undergo this procedure but asks two questions: "Can you put in the same hip replacement as my previous doctor used?" and "How much longer will my left hip last?" Unsure about the details of the previous surgery, you schedule another appointment with her in four weeks, reassuring her that you will provide more information regarding the longevity of the total hip replacement given her specific findings on examination and radiographs.

The Literature Search

To provide this patient with the most specific information about the longevity of what is eventually confirmed as a Charnley prosthesis in her left hip, one can access PubMed (a

---

This article is the second in a series designed to help the orthopaedic surgeon use the published literature in practice. In the first article in the series, we presented guidelines for making a decision about therapy and focused on randomized controlled trials. In this article, we focus on evaluating nonrandomized studies that present information about a patient's prognosis.

database of medical literature) from a computer Internet site at www.ncbi.nlm.nih.gov/PubMed. The importance of a careful search cannot be understated. Databases such as MEDLINE typically identify a small proportion of all available studies. When a particular database does not elicit an article of interest, other strategies should be employed. Additional strategies to find relevant articles include use of multiple databases (EMBASE, MEDLINE, and PubMed), review of bibliographies of articles on the topic, review of recent textbooks for relevant references, and consulation with content experts.

By entering the key words (with the Boolean operator AND) "total hip arthroplasty" AND "survival" AND "risk factors," thirty-two articles are identified in PubMed. Scanning through the titles reveals that two articles appear particularly promising: "Poor Bone Quality or Hip Structure as Risk Factors Affecting Survival of Total-Hip Arthroplasty"[1] and "Primary Hybrid Total Hip Replacement, Performed with Insertion of the Acetabular Component without Cement and a Precoat Femoral Component with Cement."[2]

Background

Why Measure Prognosis?

Surgeons help patients by diagnosing what is wrong with them, by administering treatment that does more good than harm, and by giving them an indication of the natural history of their disease or the anticipated outcome of its treatment. To achieve the second and third goals, surgeons require studies of patient prognosis—that is, investigations examining the possible outcomes of a disease or operative procedure and the probability with which they can be expected to occur. To estimate patients' prognoses, we examine outcomes in groups of patients with a similar clinical presentation—for example, patients in the first weeks after revision total hip surgery. Surgeons may then refine the prognosis by looking at subgroups and deciding the one in which their patient belongs. One may define these subgroups by demographic variables such as age (younger patients may fare better than older ones), disease-specific variables (outcomes may differ according to whether, for example, the fracture was open or closed), or comorbid factors (for example, those with underlying diabetes may fare badly). When these variables or factors accurately predict which patients will do better or worse, they are called *prognostic factors*[3].

Authors often distinguish between prognostic factors and *risk factors,* patient characteristics associated with the development of the disease in the first place. For example, low bone density is an important risk factor for the development of a hip fracture in the elderly, but it is not as important a prognostic factor in determining survival after hip fracture. The issues involved in assessing the validity of studies of prognostic factors and those of risk factors, and in using the results in patient care, are identical. One may also think of risk factors as one particular kind of prognostic factor.

Knowledge of a patient's prognosis can help surgeons to make the right diagnostic and treatment decisions. If a patient will get well anyway, the clinician should not recommend

high-risk invasive procedures or waste resources on expensive or potentially toxic treatments. If a patient is at low risk for an adverse outcome, even beneficial treatments may not be worth it, especially if the risks of treatment outweigh the benefits. In general, patients will be less willing to accept the risk of a treatment complication when the treatment is unlikely to substantially reduce their risk of a clinically important adverse outcome event. For example, in order to prevent a single event of venous thrombosis in patients undergoing a carpal tunnel release, anticoagulant prophylaxis would have to be administered to hundreds of patients because these patients are at extremely low risk for clinically important venous thrombosis[4]. In this case, the higher risk of bleeding may outweigh the benefits of anticoagulant prophylaxis. Conversely, surgeons may be reluctant to offer operations to patients who are destined to have a poor result; for example, they may not wish to perform an Ilizarov bone reconstruction in a young smoker, who has a high risk of clinically important complications (nonunion, amputation, and infection)[5].

Knowledge of prognosis is also useful for resolution of issues broader than the care of the individual patient. Organizations may attempt to compare the quality of care across health-care providers, or provider institutions, by measuring the outcomes of care. Differences in outcome may, however, be due to variability in the underlying severity of illness and not to the treatments, providers, or health-care institutions under study. If one knows patients' prognoses, one may be able to compare populations, and adjust for differences in prognosis, to obtain a more accurate indication of how treatment is affecting outcome.

Study Designs for Prognostic Studies

It is usually impossible or unethical to randomize patients to different prognostic factors. For example, it would clearly be unacceptable to randomize consecutive patients to *smoking* or to *no smoking* to determine if smoking negatively affects fracture-healing. The best study design to identify the presence of and determine the increased risk associated with a prognostic factor is a cohort study. Surgeons can conduct a cohort study by following one or more groups (cohorts) of individuals who have not yet experienced an adverse event and by monitoring the number of outcome events over time. An ideal cohort study consists of a well-defined sample of individuals representative of the population of interest and uses objective outcome criteria. A potential cohort study may document the smoking status of all consecutive patients with a tibial shaft fracture and compare rates of nonunion (or time to fracture union).

Cohort studies may be prospective in that they begin at a specified point in time (such as the time of the onset of symptoms or the time of fracture) and move forward in time to evaluate the effect of a potential prognostic factor (for example, operative compared with nonoperative treatment) on specified outcomes after a predetermined duration of follow-up. Such studies have the advantage of ensuring that all of the relevant data are collected at the start of the study, but they are

often time-consuming to conduct. Cohort studies may also be retrospective; that is, they can begin at a specified point in time and move backward in time to collect data on potential risk factors for an undesirable outcome (such as fracture nonunion) or to compare the results of two treatments. The obvious advantage of this approach is that less time is required to collect the data; however, the major drawback is the investigators' inability to ensure the quality of the collected data as they often must rely on patient records for information. In most instances, all relevant data cannot be collected because of the variability of the reporting in the hospital charts.

To study prognostic factors, surgeons can use an alternative study design in which they collect "cases" of individuals who have already had an outcome event and compare them with those of "controls" who have not. In these "case-control" studies, surgeons can count the number of individuals with each prognostic factor in both groups—for example, they can determine whether patients who had aseptic loosening of a hip replacement were more likely to have decreased bone density than those who did not. Case-control studies are limited by the retrospective nature of the data collection, with the investigators often relying on hospital charts or the patient's memory. Moreover, case-control studies do not provide information about the absolute risk of an adverse event; they can only demonstrate the relative odds[3]. Despite these limitations, case-control studies can be useful when the outcome of interest is very rare or the duration of follow-up needed to detect the outcome of interest is long.

Are the Results of the Study Valid?

Primary Guides (Step 1)

Was there a representative sample of patients?

A prognostic study is biased if it yields a systematic overestimate or underestimate of the likelihood of adverse outcomes in the patients under study. When a sample is systematically different from the underlying population, and is therefore likely to be biased because patients will have a better or worse prognosis than those in the underlying population, that sample is labeled as *unrepresentative*.

How can surgeons recognize an unrepresentative sample? First, they can look to see if patients pass through some sort of "filter" before entering the study. If they do, the sample is likely to be systematically different from the underlying population. One such filter is the sequence of referrals that leads patients from primary to tertiary centers. Tertiary centers often care for patients with rare disorders or more severe illness. Research describing the outcomes of patients in tertiary centers may not be applicable to the general patient with the disorder. For example, intensive-care physicians at university-based units are more likely to withdraw life support (ventilators) than are physicians based in the community[6]. This is likely a result of the severity of injuries seen in patients treated in tertiary care hospitals. When an individual is admitted to a hospital with a head injury, family members will want to know the risk of death, but studies of mortality from head injury are highly variable[7]. Patients with an isolated head injury, who are

often treated in community centers, have a 14% rate of mortality, whereas those who present to tertiary care centers (level-I trauma centers) have been reported to have a 46% mortality rate[7]. This is likely due to the severity of head injury (Glasgow coma scale score <9) as well as associated injuries (fractures and injuries of abdominal organs) in patients who are transferred to a trauma center.

Failure to clearly define the patients who entered the study increases the risk that the sample will be unrepresentative. To help determine the representativeness of the sample, look for a clear description of which patients were included and excluded from a study. How the sample was selected and the objective criteria used to diagnose the disorder should both be clearly specified.

Were the patients sufficiently homogeneous with respect to prognostic risk? If not, did the investigators provide estimates for all clinically relevant subgroups?

Prognostic studies are most useful if all patients in the entire group are similar enough for the outcome of the group to be applicable to each member. This will be true only if patients are at a similar, well-described point in their disease process. The point in the clinical course need not be early, but it does need to be consistent. In surgical studies, one might decide to describe patients at the time of an operative procedure such as a joint arthroplasty or fracture fixation. It is important for readers to be sure that the patients undergoing these surgical procedures are similar—that is, that the stage of disease is relatively constant.

Assuming that this is the case, it is important to consider other factors that might influence patient outcome. Consider an example of total hip arthroplasty. A study examining the survival rate of hip prostheses that pools patients with rheumatoid arthritis and osteoarthritis without distinguishing between them may not be very useful if these two groups have different prognoses. Furthermore, if the overall mortality reported in a study is 50%, but the patient population is made up of two identifiable subgroups, one with a mortality rate near zero and the other with a mortality rate near 100%, the 50% estimate will be valid for the whole group but not for any individual in that group. If the patients are heterogeneous with respect to the risk of an adverse outcome, the study will be much more useful if the investigators define the two subgroups at lower and higher risk than the overall group.

Pincus et al. followed a cohort of patients with rheumatoid arthritis for fifteen years[8]. They separated the patients into a number of cohorts depending on their demographic characteristics, disease variables, and functional status. They found that older patients and those with greater impairment of functional status (for example, slower walking time and problems in activities of daily living) died earlier than the others. In another study, Kitamura et al. evaluated the outcomes following hip fractures in 1217 patients[9]. They identified an increased risk of mortality for patients greater than eighty years old, those with dementia, those of male gender, and those with a history of a hip fracture[9].

USER'S GUIDE TO THE ORTHOPAEDIC LITERATURE:
HOW TO USE AN ARTICLE ABOUT PROGNOSIS

Investigators not only must consider all important prognostic factors, but they also must consider them in relation to one another. Consider a study by Zuckerman et al., who examined risk factors for mortality following hip fractures[10]. They identified an operative delay of three or more days as a significant predictor of mortality (p = 0.04). Taken at face value, this would suggest that if surgeons could avoid a delay they might reduce the mortality rate. However, to properly understand the impact of delay in operative treatment, one must simultaneously consider other prognostic features, such as the severity of preexisting medical conditions. In assessing the importance of operative delay, the investigators must separately examine the relative risk of mortality in patients with and without severe medical conditions in two groups: those in whom operative treatment was delayed and those in whom it was not. This separate consideration is called an *adjusted analysis.* Once adjustments were made for severity of preexisting medical conditions (American Society of Anesthesiologists grades I, II, and III), Zuckerman et al. found that operative delay no longer predicted the risk of mortality. It turned out that patients in whom operative treatment was delayed were sicker than those who underwent the operation earlier. It was the underlying severity of illness, not the operative delay, that was responsible for the increased mortality.

If there are a few variables that have a major impact on prognosis, investigators may use a simple technique called *stratified analysis.* This can be accomplished by dividing patients into groups, or strata, on the basis of their prognosis (for example, diabetics and nondiabetics) and evaluating outcomes separately for each stratum. If there is a large number of variables that have a major impact on prognosis, the investigators should use sophisticated statistical techniques (multiple regression or logistic regression) to determine the most powerful predictors. Such an analysis may lead to a clinical prediction rule that guides clinicians in simultaneously considering all of the important prognostic factors.

As surgeons, we are often interested in prediction. We want to know which person will have an outcome of interest (such as mortality) and which person will not as well as which patient will do well and which patient will do poorly. Regression techniques are useful in addressing this sort of question. Generally, when we construct regression equations, we refer to the predictor variable (independent variable) as x and the target variable (dependent variable) as y. A simple regression equation may read as follows: Y (loosening) = K (constant) + B (patient age), where B is the slope of the best-fit regression line and K is the y-intercept. If there is only one variable, the analysis is referred to as univariable (or simple) regression analysis. If there are multiple predictor variables (for example, patient age, type of arthritis, severity of arthritic condition, activity level, weight, cementing techniques, and acetabular or femoral stem orientation), then the regression analysis is labeled multivariable. The target, or dependent variable, can be dichotomous (for example, mortality or hip revision) or continuous (for example, time to revision surgery). When dichotomous target variables are utilized, it is referred to as *logistic regression analysis.*

Lee et al. developed such a prediction rule to estimate the risk of cardiac complications in patients undergoing noncardiac surgery[11]. This so-called revised cardiac risk index was derived from a cohort of 4315 patients who were undergoing elective noncardiac surgery. Using sophisticated statistical regression techniques, these authors identified six variables (each given 1 point if present) that proved to be important predictors of cardiac complications. These included high-risk surgery (such as intrathoracic, suprainguinal vascular, or intraperitoneal surgery), coronary artery disease, congestive heart failure, a history of cerebrovascular disease, insulin treatment for diabetes mellitus, and a preoperative serum creatinine level >2.0 mg/dL (>177 μmol/L). This prediction rule was validated in a separate cohort of 1422 patients undergoing elective noncardiac surgery. Patients with no risk factors had a 0.5% prevalence of cardiac complications, whereas those with one, two, or three or more risk factors had a 1.3%, 3.6%, and 9.1% prevalence of cardiac complications, respectively.

In another example, Signorini et al. used multivariate logistic regression to derive a model, ultimately consisting of five variables, to predict the one-year survival rate in a group of 372 patients with traumatic brain injury who presented to a trauma unit in Edinburgh, Scotland[12]. These five variables included age, Glasgow coma scale score, injury severity score, pupillary reaction, and evidence of a hematoma on a computed tomography scan. Those authors validated their prediction rule in a separate cohort of 520 patients.

How can surgeons decide if a group is sufficiently homogeneous with respect to risk? On the basis of one's clinical experience and one's understanding of the biological characteristics of the condition being studied, can one think of factors that the investigators have neglected that are likely to define subgroups with very different prognoses? To the extent that the answer is yes, the validity of the study is compromised. For instance, readers of a report on predictors of a reoperation following fracture fixation will find the results less compelling if the investigators failed to examine the influence of fracture severity.

Secondary Guides (Step 2)
Was follow-up sufficiently complete?
As with randomized trials, a high patient dropout rate also threatens the validity of a cohort study of prognosis. As the number of patients who do not return for follow-up increases, the likelihood of bias increases as well because those who are followed may be at systematically higher or lower risk than those who are not followed. What proportion of patients lost to follow-up seriously threatens a study's validity? The answer depends on the relationship between the proportion of patients who are lost and the proportion of patients who had the adverse outcome of interest. The larger the number of patients whose fate is unknown relative to the number who had an event, the greater the threat to the study's validity.

For instance, let us assume that 30% of a particularly high-risk group (such as elderly patients with renal failure[13])

TABLE I User's Guide to the Surgical Literature: Guide to an Article About Prognosis
I. Guides for validity (study methodology)
Step 1: Assess primary guides:
Was there a representative sample of patients?
Were the patients sufficiently homogeneous with respect to prognostic risk? If not, did investigators provide estimates for all clinically relevant subgroups?
Step 2: Assess secondary guides:
Was follow-up sufficiently complete?
Were objective and unbiased outcome criteria used?
II. Understanding the study results
Step 3: How likely are the outcomes to occur over time?
Step 4: How precise are the estimates of likelihood?
III. Using the results to determine patient care (applying the results to your patient)
Step 5: Were the study patients and their management similar to your own?
Step 6: Was the follow-up sufficiently long?
Step 7: Can you use the results to determine the management of your patient?

have had an adverse outcome (such as implant loosening) in a long-term study of the results of hip arthroplasty. If 10% of the patients were lost to follow-up, the true rate of patients with a loose prosthesis may be as low as 27% or as high as 40%. Across this range, the clinical implications would not change appreciably and the loss to follow-up does not threaten the validity of the study. However, in a much lower-risk patient sample (otherwise healthy, active women, for example), the observed loosening rate may be 1%. In this case, if one assumed that all of the patients lost to follow-up (10% of the group) had a loose prosthesis, the event rate of 11% might have very different implications.

A large loss to follow-up constitutes a more serious threat to validity when the patients who were lost may differ from those who were easier to find. For example, after much effort, 180 of 186 patients treated for neurosis were followed in one study[14]. The death rate was 3% among the three-fifths who were easily traced, but it was 27% among those who were more difficult to find. If it is plausible that the fate of those who were followed differed from the fate of those who were lost (and it is in most prognostic studies), a loss to follow-up that is large in relation to the proportion of patients with the adverse outcome of interest constitutes an important threat to validity.

Were objective and unbiased outcome criteria used?

Outcome events can range from those that are objective and easily measured (death), to those that require some judgment (healing of a fracture), to those that require considerable judgment and are challenging to measure (disability or quality of life). Investigators should make every attempt to identify previously validated and reliable scales when contemplating the assessment of quality of life or functional status. Investigators should clearly define their target outcomes before the study and, whenever possible, base their criteria on the most clinically relevant measures. In addition, investigators should specify the intensity and frequency of monitoring (active follow-up). As the subjectivity of the outcome definition increases, it becomes more important that individuals determining the outcomes are blinded to the presence of prognostic factors.

In an observational study of thirty-four patients treated with core decompression for nontraumatic osteonecrosis of the femoral head, researchers evalu-

ated patient outcomes at a mean of ten years[15]. They classified patients according to the radiographic stage of the disease as well as risk factors predisposing to osteonecrosis (corticosteroid use, excessive alcohol intake, adrenocorticotropic hormone treatment, or idiopathic osteonecrosis). At the time of follow-up, outcome assessors unblinded to prognostic factors categorized the outcomes of the core decompressions as successful (no symptoms or radiographic progression) or as a failure (either radiographic or clinical). Because it was relatively subjective, the decision about a successful outcome in this situation may have been influenced by prior knowledge of prognostic factors for disease progression.

Applying Validity Criteria to Survivorship Studies of Total Hip Arthroplasty

Table I lists the key criteria for ensuring the validity of a study of prognosis. We can apply these criteria to the articles that we found that addressed the patient scenario presented at the beginning of this article. Recall that our literature search revealed two relevant articles. Answers to the questions in Table I may not always be reported by authors. In such cases, the reader has two options: assume that if an item was not reported it was not addressed or assume that the item was addressed but it was not reported because of an oversight on the part of the authors. If the latter approach is chosen, the reader should attempt to correspond with the primary author. The urgency with which a response is required will often dictate the choice of communication (telephone call or written correspondence).

In the first article identified[1], 411 patients with advanced hip disease underwent a total hip arthroplasty between 1972 and 1988. One surgeon with training in the procedure as initially performed by Charnley carried out the operations at a university hospital in Japan. All patients were identified at the time of the operative procedure, so we cannot be sure that the patients were at the same stage of disease. A better time to identify patients may have been at the onset of the arthritis.

A number of factors that might influence the risk of aseptic loosening include patient age or weight, type of arthritis, severity of the arthritic condition, activity level, cementing techniques, and acetabular or femoral stem orientation. As will be seen, the investigators tested all of these factors.

The investigators excluded six patients in whom deep infection developed, and they followed 100% of the remainder; thus, 405 of the 411 patients were followed, for a mean of 14.1 years (range, one month to twenty-six years). The au-

thors provided a detailed definition of failure of radiographic fixation (loosening) and revision surgery (a more objective outcome measure). An outcome assessor independent of the surgeon who performed the operations evaluated patient outcomes[16]. The outcome assessor was not blinded to potentially important prognostic variables[16]. Thus, the sample is likely representative of Japanese patients with advanced osteoarthritis who present to university settings for primary total hip arthroplasty, the investigators identified all relevant
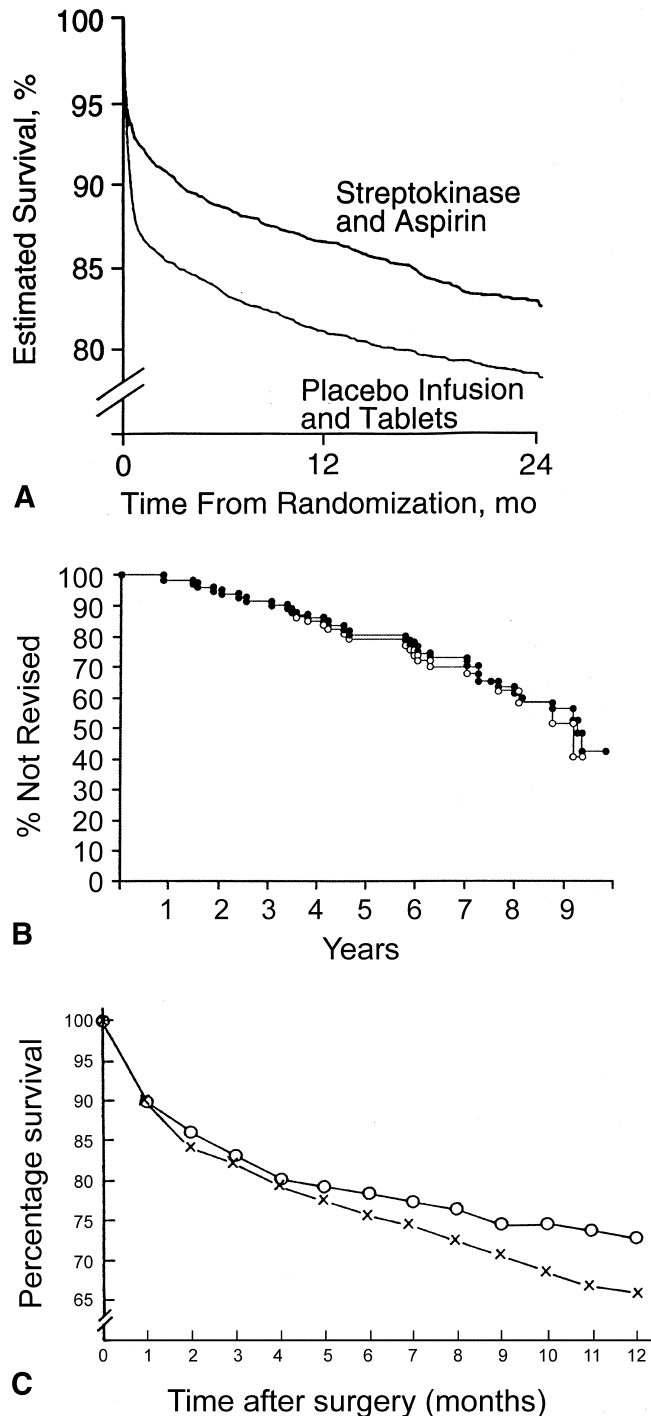

Fig. 1

A: Survival after myocardial infarction of patients treated with streptokinase and aspirin compared with those treated with a placebo. (Reproduced, with modification, from: ISIS-2 [Second International Study of Infarct Survival] Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. Lancet. 1988;2:349-60. Reprinted with permission.) B: Need for revision after total hip arthroplasty in two cohorts of patients treated in the same center. (Reproduced, with modification, from: Dorey F, Amstutz HC. The validity of survivorship analysis in total joint arthroplasty. J Bone Joint Surg Am. 1989;71:544-8.) C: Survival after hip fracture. (Reproduced, with modification, from: Bredahl C, Nyholm B, Hindsholm KB, Mortensen JS, Olesen AS. Mortality after hip fracture: results of operation within 12 h of admission. Injury. 1992;23:83-6. Reprinted with permission.)

USER'S GUIDE TO THE ORTHOPAEDIC LITERATURE:
HOW TO USE AN ARTICLE ABOUT PROGNOSIS

prognostic factors, follow-up was excellent, and the outcome measures were objective. With the major criteria satisfied, lack of blinding to prognostic features does not constitute a major threat to validity.

In the second article, Clohisy and Harris[2] followed 107 patients in whom 121 primary hybrid total hip replacements had been performed between 1984 and 1987. The operations were conducted at a university hospital in the United States by one surgeon. All patients were identified at the time of the operative procedure. The investigators collected information on the following potentially important variables: the reason for the hip surgery, type of acetabular component, acetabular preparation, and femoral preparation.

Eighty-six patients with 100 total hip arthroplasties were followed for a mean of ten years. None of the hips in the fifteen patients who died and the six who were lost to follow-up had required revision surgery at a mean of 3.2 years. The investigators provided detailed descriptions of the operative procedure and their definitions of femoral osteolysis, acetabular osteolysis, and polyethylene wear. The outcome measures were evaluated by an independent orthopaedic surgeon. However, it was not reported whether the orthopaedic surgeon was blinded to prognostic factors. Thus, in this study, the authors recruited a representative sample but failed to examine the potential impact of prognostic features, were moderately successful with regard to following patients, and used objective outcome criteria. Again, because of their objectivity, judgments about whether outcomes had occurred are unlikely to have been influenced by the absence of blinding.

Results

Having decided that a study's methods suggest that it will yield valid results, readers should be aware of common strategies to relay information about a study of prognosis.

How likely are the outcomes to occur over time? (Step 3)
The quantitative results from studies of prognosis or risk are the numbers of events that occur over time. We will use the example of a man asking a physician about the prognosis for his elderly father with a hip fracture to illustrate common expressions that provide complementary information about prognosis.

The patient's son asks: "What are the chances that my father will still be alive in two years?" A high-validity study of the prognosis for patients with a hip fracture[17] provides a simple and direct answer in absolute terms. Two years after hip the surgery, about 25% of the patients had died. Thus, there is about a one-in-four chance that the father will die in the next two years.

The patient's son might then tell the physician that the only person whom he knows with a previous hip fracture is a sixty-five-year-old aunt who had the fracture fixed almost ten years ago and is still living. He is surprised that his father's chance of dying in the next two years is so high. This gives
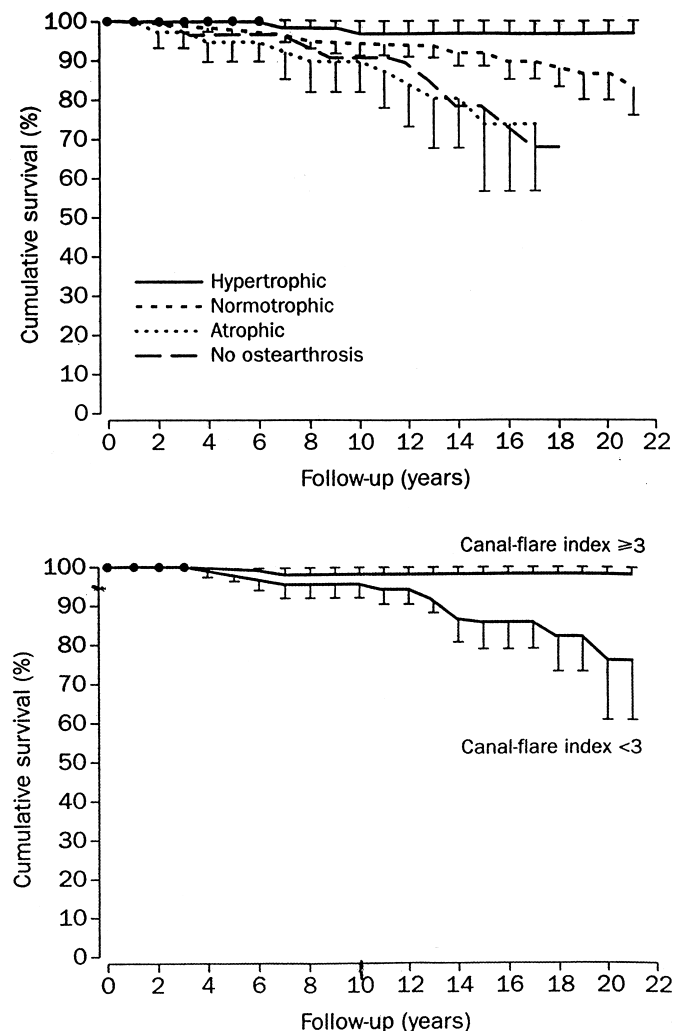

Fig. 2
Implant survival based on type of arthritis and canal-flare index. (Reprinted, with permission, from: Kobayashi S, Saito N, Horiuchi H, Iorio R, Takaoka K. Poor bone quality or hip structure as risk factors affecting survival of total-hip arthroplasty. Lancet. 2000;355:1499-504.)

the surgeon the opportunity to discuss some of the prognostic factors for death of patients with a hip fracture. The just-mentioned study[17] suggested that older patients, those with more severe dementia, and men were more likely to die than were those without these characteristics.

The son might then ask whether his father's chances of survival will change with time—that is, might the risk of death be relatively low for the next two years and then jump sharply after that? Neither the absolute nor the relative expressions of the results address this question. For this answer, we should turn to a survival curve, which is a graph of the number of events over time (or, conversely, the chance of the patient being free of those events over time). The events must be discrete (for example, death, revision surgery, and complications), and the precise time at which they occur

USER'S GUIDE TO THE ORTHOPAEDIC LITERATURE:
HOW TO USE AN ARTICLE ABOUT PROGNOSIS

must be known. Figure 1 shows three survival curves, one showing survival after myocardial infarction (Panel A)[18], one showing the results of hip replacement (Panel B)[19], and the third showing survival after hip fracture (Panel C)[20]. Note that the chance of dying after a myocardial infarction is highest shortly after the event (reflected by an initially steep slope of the curve, which then flattens), whereas very few hip replacements require revision until much later (this curve starts out flat and then steepens). The survival curve for patients with a hip fracture suggests that the risk of dying increases at a steady rate after the operation.

How precise are the estimates of likelihood? (Step 4)

The more precise the estimate of a prognosis, the less the uncertainty regarding the estimated prognosis and the more useful it is. Usually, risks of adverse outcomes are reported with their associated 95% confidence intervals. The 95% confidence interval defines the range of risks within which (if the study was valid) it is highly likely that the true risk lies. For example, if the 95% confidence interval for the risk of radiographic loosening following hip arthroplasty is 5% to 10%, then readers can be assured (assuming that they believe that the study is valid) that the true risk lies somewhere between 5% and 10%. Put another way, if the study were repeated 100 times, the rate of radiographic loosening would be between 5% and 10% ninety-five of those 100 times. Note that, in most survival curves, the results are derived from more patients during the earlier follow-up periods than during the later periods (as a result of losses to follow-up and because patients are not enrolled in the study at the same time). This means that the survival curves are more precise in the earlier periods, indicated by narrower confidence bands around the left part of the curve. For instance, the 95% confidence intervals in the study of prognosis after hip replacement by Kobayashi et al.[1] are narrow in the first ten years following the surgery and widen after twenty years, as fewer patients remain without an event (Fig. 2).

Applying Results Criteria to Survivorship Studies of Total Hip Arthroplasty

The study by Kobayashi et al.[1] showed that the risks of radiographic failure and revision in the first ten years after arthroplasty were 6% (95% confidence interval, 3.8% to 8.7%) and 1% (95% confidence interval, 0% to 2.3%), respectively. At twenty years, these values were 16% (95% confidence interval, 10.7% to 21.1%) and 10% (95% confidence interval, 4.6% to 14.9%), respectively.

The investigators examined whether seven patient variables (gender, age, diagnosis, Charnley functional category, postoperative activity, height, and weight), four radiographic variables (polyethylene wear rate, implant orientation, canal-flare index, and biological classification of osteoarthritis), and three surgical variables (cementing, implant design, and preparation of the acetabulum) predicted aseptic loosening after total hip arthroplasty. Of these factors, rapid polyethyl-

ene wear and the classification of the osteoarthritis (hypertrophic, normotrophic, or atrophic) significantly predicted revision of the acetabular component, and a low canal-flare index (<3) predicted loosening of the femoral component (Fig. 2). However, there have been concerns in the literature regarding the use of terminology such as "hypertrophic osteoarthritis."[21] Accordingly, it may not be a helpful predictor in this situation.

In the article by Clohisy and Harris[2], the risk of failure of a hybrid total hip replacement was 4% (95% confidence interval, 2% to 7%) at ten years. However, the authors did not adjust the estimates of survival for important prognostic factors. Thus, the summary estimate represents one from a heterogenous group of patients. Moreover, if we assume that all six patients lost to follow-up (5.6% of the original series) had a failure of the total hip arthroplasty, then the risk of failure may be as high as 9.6%.

Applicability

Were the study patients and their management similar to your own? (Step 5)

The authors should describe the study patients in enough detail so that you can compare them with your patients. This should include not only the patients' characteristics but also how those characteristics were defined. One factor that could strongly influence outcome but is rarely reported in prognostic studies is therapy. Therapeutic strategies often vary markedly among institutions and change over time as new treatments become available or old treatments regain popularity. To the extent that our interventions are therapeutic or detrimental could determine whether overall patient outcome improves or worsens. For example, while skeletal traction was the most common definitive treatment of femoral shaft fractures for decades before the 1970s, intramedullary nails have long since become the standard of care. Studies that fail to provide sufficient details about the therapeutic strategies limit the reader's ability to assess the applicability of the results of the study to his or her own patients. The issue of evolving therapy is even more relevant in long-term outcome studies of arthroplasty. Over a ten to twenty-year time span, new implants, technologies, and modifications of surgical technique will limit the ability of investigators to report on a single series of patients treated with the "exact" same implants. Such studies must provide details regarding the various types of implants and surgical modifications over the years so that readers can assess the generalizability of the results to their own patients.

Was the follow-up sufficiently long? (Step 6)

Since illness often precedes the development of an outcome event by a long period, investigators must follow patients for long enough to detect the outcomes of interest. This is particularly true if your patient is interested in his or her risk over a long period of time. A study in which patients were followed for five years after hip replacement would be of little use.

Can you use the results to determine the management of your patient? (Step 7)

Prognostic data often provide the basis for sensible decisions about therapy. Knowing the expected clinical course of your patient's condition can help you to judge whether treatment should be offered at all. For example, anticoagulant prophylaxis following hip surgery markedly decreases the risk of proximal venous thrombosis in patients with a history of thromboembolism or malignant disease (risk of proximal venous thrombosis in category-III patients, 10% to 20%) and is indicated for all such patients with these risk factors[4]. However, otherwise active, young patients (less than forty years old) with uncomplicated surgery are at low risk for proximal venous thrombosis (0.4%). While anticoagulant therapy will have the same effect in both high and low-risk groups, most patients at low risk (0.4%) are likely to think that, for them, the risk of anticoagulant therapy (1% to 2% risk of bleeding) outweighs the benefit.

Even if knowledge of the prognosis does not help the physican choose a therapy, it can help him or her counsel a concerned patient or relative. Some conditions, such as asymptomatic hiatal hernia or asymptomatic colonic diverticula, have such a good overall prognosis that they have been termed nondisease[22]. On the other end of the spectrum, a uniformly bad prognosis provides the clinician with a starting place for a discussion with the patient and family, leading to counseling about end-of-life concerns.

Resolution of the Scenario

Having addressed issues regarding study results and applicability, you can now apply these criteria to the two eligible studies.

Review of the validity criteria suggests that Kobayashi et al.[1] obtained an unbiased assessment of risk in their cohort. The patients in their study were mainly women (89%), were an average of sixty years old (range, twenty-eight to eighty-five years old), and weighed an average of 62 kg[16]. Your patient resembles the majority of those in the cohort in terms of age, gender, and body habitus. Patients were followed for up to twenty-six years, allowing the investigators to provide estimates for patients up to twenty years after the operation. Thus, you can readily generalize the results to your patient's care and provide her with an estimate of her long-term prognosis, with one caveat: do you believe that your surgical skills are similar to those of the surgeons in the study? The study's presentation of one surgeon's experience at an academic center may raise concerns about its generalizability to a surgical practice with much less volume.

Given that this woman appears to have only a single risk factor for aseptic loosening of the femur (a narrow canal-flare index), you can be reasonably confident that, on the basis of the survival curve (Fig. 2), she has a 2% to 3% risk of femoral loosening over ten years and just slightly more than a 20% risk over twenty years. It has now been twelve years since her left total hip replacement. She can be assured that she has at least an 80% chance of not needing another operation in the next eight years. To be most accurate, you should employ conditional probability formulas using a Bayesian approach. Briefly, Bayes' theorem uses new information (conditional probability) to update old information about the probability of an event. Clearly, while the overall risk of implant loosening is 20% over twenty years, a woman in whom the replacement has already survived for twelve years has a different probability of loosening over the next eight years. Thus, given the fact that the prosthesis has already survived twelve years without revision surgery in this patient, her risk of having a reoperation is 6% (97% survival at ten years minus 80% survival at twenty years equals a difference of 17%—the chance for prosthetic survival is now one in sixteen, or $10/160 \times 100 = 6\%$).

The study by Clohisy and Harris[2] raises questions concerning the loss to follow-up and failed to specify how prognostic factors may influence outcome. Moreover, you are unsure whether your patient is similar to those in this article. For these reasons, the results from this article may not be applicable to your patient.

At your four-week follow-up visit with this patient, you carefully explain the surgical procedure and the risk factors for loosening in the future. She is pleased to know that her chances of having a well-fixed hip that is not painful are upward of 94% for ten years.

Conclusion

We have presented an approach to critical appraisal of a study describing important outcomes following total hip arthroplasty along with the frequency with which they can be expected to occur. Authors of studies of prognostic factors can limit bias by selecting patients at a similar point in the course in their disease, ensuring completeness of follow-up, providing separate estimates for different prognostic groups, and utilizing unbiased and objective outcomes. ■

Mohit Bhandari, MD, MSc
Gordon H. Guyatt, MD, MSc
Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Center, Room 2C12, 1200 Main Street West, Hamilton, ON L8N 3Z5, Canada. E-mail address for M. Bhandari: bhandari@sympatico.ca

Marc F. Swiontkowski, MD
Department of Orthopaedic Surgery, University of Minnesota, Box 492, Delaware Street N.E., Minneapolis, MN 55455

References

1. **Kobayashi S, Saito N, Horiuchi H, Iorio R, Takaoka K.** Poor bone quality or hip structure as risk factors affecting survival of total-hip arthroplasty. *Lancet.* 2000;355:1499-504.

2. **Clohisy JC, Harris WH.** Primary hybrid total hip replacement, performed with insertion of the acetabular component without cement and a precoat femoral component with cement. An average ten-year follow-up study. *J Bone Joint Surg Am.* 1999;81:247-55.

3. **Laupacis A, Wells G, Richardson WS, Tugwell P** for the **Evidenced-Based Medicine Working Group.** Users' guides to the medical literature. V. How to use an article about prognosis. *JAMA.* 1994;272:234-7.

4. **Ginsberg J, Hirsh J.** *Management guidelines in venous thromboembolism.* 2nd ed. Hamilton, ON: BC Decker; 1999. p 42-53.

5. **Marsh D, Shah S, Elliott J, Kurdy N.** The Ilizarov method in nonunion, malunion and infection of fractures. *J Bone Joint Surg Br.* 1997;79:273-9.

6. **Bach PB, Carson SS, Leff A.** Outcomes and resource utilization for patients with prolonged critical illness managed by university-based or community-based subspecialists. *Am J Respir Crit Care Med.* 1998;158:1410-5.

7. **Shedden PM, Moulton RJ, Sullivan I, Hotz G, Tucker WS, Muller PJ.** Effect of population characteristics on head injury mortality. *Pediatr Neurosurg.* 1990;16:203-7.

8. **Pincus T, Brooks RH, Callahan LF.** Prediction of long-term mortality in patients with rheumatoid arthritis according to simple questionnaire and joint count measures. *Ann Intern Med.* 1994;120:26-34.

9. **Kitamura S, Hasegawa Y, Suzuki S, Sasaki R, Iwata H, Wingstrand H, Thorngren KG.** Functional outcome after hip fracture in Japan. *Clin Orthop.* 1998;348:29-36.

10. **Zuckerman JD, Skovron ML, Koval KJ, Aharonoff G, Frankel VH.** Postoperative complications and mortality associated with operative delay in older patients who have a fracture of the hip. *J Bone Joint Surg Am.* 1995;77:1551-6.

11. **Lee TH, Marcantonio ER, Mangione CM, Thomas EJ, Polanczyk CA, Cook EF, Sugarbaker DJ, Donaldson MC, Poss R, Ho KK, Ludwig LE, Pedan A, Goldman L.** Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation.* 1999;100:1043-9.

12. **Signorini DF, Andrews PJ, Jones PA, Wardlaw JM, Miller JD.** Predicting survival using simple clinical variables: a case study in traumatic brain injury. *J Neurol Neurosurg Psychiatry.* 1999;66:20-5.

13. **Devlin VJ, Einhorn TA, Gordon SL, Alvarez EV, Butt KM.** Total hip arthroplasty after renal transplantation. Long term follow-up study and assessment of metabolic bone status. *J Arthroplasty.* 1988;3:205-13.

14. **Sims AC.** Importance of a high tracing-rate in long-term medical follow-up studies. *Lancet.* 1973;2:433-5.

15. **Bozic KJ, Zurakowski D, Thornhill TS.** Survivorship analysis of hips treated with core decompression for nontraumatic osteonecrosis of the femoral head. *J Bone Joint Surg Am.* 1999;81:200-9.

16. **Kobayashi S.** Personal communication.

17. **Kenzora J, Magaziner J, Hudson J, Hebel JR, Young Y, Hawkes W, Felsenthal G, Zimmerman SI, Provenzano G.** Outcome after hemiarthroplasty for femoral neck fractures in the elderly. *Clin Orthop.* 1998;348:51-8.

18. **ISIS-2 (Second International Study of Infarct Survival) Collaborative Group.** Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet.* 1988;2:349-60.

19. **Dorey F, Amstutz HC.** The validity of survivorship analysis in total joint arthroplasty. *J Bone Joint Surg Am.* 1989;71:544-8.

20. **Bredahl C, Nyholm B, Hindsholm KB, Mortensen JS, Olesen AS.** Mortality after hip fracture: results of operation within 12 h of admission. *Injury.* 1992;23:83-6.

21. **Jones A, Doherty M.** ABC of rheumatology. Osteoarthritis. *BMJ.* 1995;310:457-60.

22. **Meador CK.** The art and science of nondisease. *N Engl J Med.* 1965;272:92-9.